PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 1/04, 14/00, C12Q 1/68

A1

(11) International Publication Number: **WO 00/02899**

(43) International Publication Date: 20 January 2000 (20.01.00)

(21) International Application Number: PCT/US99/14969

(22) International Filing Date: 1 July 1999 (01.07.99)

(30) Priority Data:
09/113,211 9 July 1998 (09.07.98) US

(71) Applicant: BIOCEPT, INC. [US/US]; Suite K, 2075 Corte del Nagal, Carlsbad, CA 92009 (US).

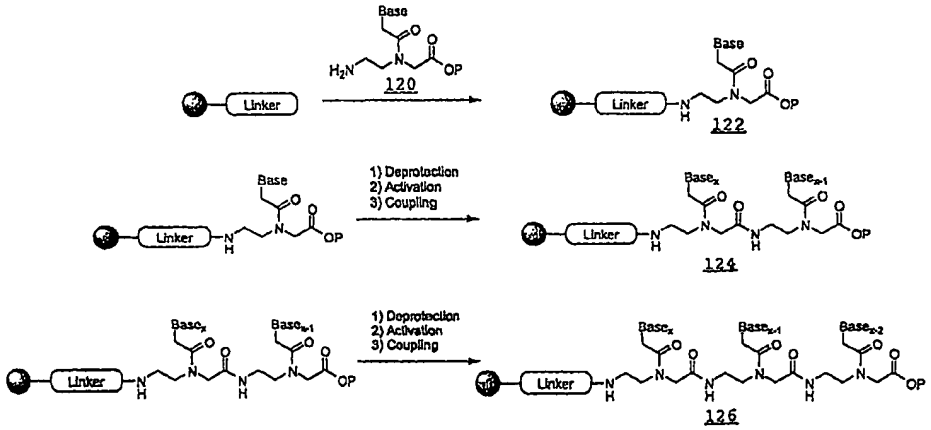
(72) Inventors: HAHN, SoonKap; 511 Via Delfin, San Clemente, CA 92672 (US). PATRON, Andrew; Number 72, 4130 Porte de Merano, San Diego, CA 92122 (US). FAGNANI, Roberto; 8972 Caminito Fresco, La Jolla, CA 92037 (US).

(74) Agents: SAMPLES, Kenneth, H. et al.; Fitch, Even, Tabin & Flannery, Suite 1600, 120 South LaSalle Street, Chicago, IL 60603-3406 (US).

(81) Designated States: AU, CA, JP, KR, SG, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD OF USING AN IMPROVED PEPTIDE NUCLEIC ACID UNIVERSAL LIBRARY TO OPTIMIZE DNA SEQUENCE HYBRIDATION



(57) Abstract

Provided herein is a system, comprising three key integrated components, which permits the rapid identification of peptide nucleic acid (PNA) oligomers useful as therapeutics, diagnostics and/or gene characterization tools. The first component is a universal PNA library that may be easily and efficiently synthesized and that most preferably incorporates one or more universal nucleotide bases into carefully selected positions within each oligomer species of the library thereby providing the library with the screening ability of a larger library without requiring synthesis of a corresponding large number of individual species. The second component in this system is a high throughput screening system that includes a number of assays designed to provide information on the binding activities of the different sequence PNAs to the target nucleotide sequence (generally, a DNA or RNA sequence). The third component of the present invention is a software system especially designed to provide rapid analysis of the data collected from the high throughput screening system and to determine therefrom the sequence base identities and sequence lengths of PNA oligomers most likely to bind to and appropriately affect the target nucleotide molecule.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

METHOD OF USING AN IMPROVED PEPTIDE NUCLEUC ACID UNIVERSAL LIBRARY TO OPTIMIZE DNA SEQUENCE HYBRIDATION

BACKGROUND OF THE INVENTION

Reagents that selectively bind to DNA or RNA are of significant interest in molecular biology and medicinal chemistry as they may be developed into gene-targeted drugs for diagnostic and therapeutic applications and may be used as tools for sequence-specific modification of DNA. To date, research directed at identifying such reagents has focused primarily upon development of various oligonucleotides and their close analogs having modified backbones, such as, phosphorothioate or methyl phosphonate backbones instead of the natural phosphodiester backbone. These reagents, however, have been found to have serious short comings, especially with respect to stability against biological degradation, solubility, cellular uptake properties and ease of synthesis. For these reasons, alternative concepts for oligonucleotide mimics have been attracting interest.

Peptide nucleic acids (PNAs) are a recently developed class of oligonucleotide mimics wherein the entire deoxyribose phosphate backbone has been replaced by a chemically different, structurally homomorphous backbone composed of (2-aminoethyl)glycine units. Despite this dramatic change in chemical makeup, PNAs recognize complementary DNA and RNA by Watson-Crick base pairing. Furthermore, PNAs have been shown to have numerous advantages over DNA and RNA oligomers. For example, PNAs lack 3′ to 5′ polarity and thus can bind in either a parallel or an antiparallel orientation to DNA or RNA. (Egholm, M. et al., Nature 365:566, 1993). It has been demonstrated that PNAs can bind double stranded DNA by invading the DNA duplex and displacing one strand to form a stable D-loop structure (Peffer et al., Proc. Natl. Acad. Sci. USA 90:10648, 1993). A further

advantage of PNAs is that they are less susceptible to enzymatic degradation (Demidov et al. Biochem. Pharmacol. 48:1310, 1994) and bind RNA with higher affinity than analogous DNA oligomers (Norton et al. Nature Biotechnology 14:615, 1996). Quite advantageously, selective hybridization of PNA to DNA is less tolerant of base pair mismatches than DNA-DNA hybridization. For example, a single base mismatch within a 16 bp PNA-DNA duplex can reduce the T_m by up to 15° C, compared to 10° C in the case of a 16 bp DNA-DNA duplex (Egholm, M. et al. Nature 365:566, 1993). Finally, in at least one example, a PNA molecule has been shown capable of mimicing a transcription factor and acting as a promoter, thus demonstrating the potential use of PNAs as gene-specific activating agents (Mollegaard et al. Proc Natl Acad Sci USA 91:3892, 1994).

The success of an oligonucleotide analog as an antisense drug requires that the oligonucleotide be taken up by cells in large enough quantities to reach its target at a concentration sufficient to cause the desired effect. Until recently PNAs have shown low phospholipid membrane permeability (Wittung et al. FEBS Letters 365:27, 1995) and have been reported to be taken up by cells very poorly (Hanvey et al. Science 258:1481, 1992; Nielsen et al. Bioconjugate Chem. 5:3, 1994; Bonham et al. Nucleic Acid Res. 23:1197, 1995), initially suggesting their potential use as anti-gene and anti-sense agents would be quite limited.

Strategies to improve the cellular uptake of PNAs by conjugating the PNA sequence to a carrier molecule have met with some limited success (Basu et al. Bioconjugate Chem. 8:481, 1997). Conjugation of PNA molecules to receptor ligand molecules has increased cellular uptake of the PNA; however, the ability of these

receptor ligand-conjugated PNA oligomers to influence biological activity once inside the target cells remains unproven. Further, using such a conjugation strategy permits the PNA oligomers to enter only those cells expressing the particular targeted receptor. Thus, an appropriate ligand molecule would have to be designed for each cell type of interest.

However, recently it has been discovered that unconjugated (aka "naked") PNA oligomers administered extracellularly can both cross cell membranes (Gray, G.D. Biochem. Pharmacol. 53:1465, 1997) and elicit a sequence-specific biological response in living cells (Richelson, E. FEBS Letters 421:280, 1998). Thus, PNAs possess the following characteristics suggesting they are well suited as therapeutic and diagnostic candidates: Cell permeability in vivo; higher specificity and stronger binding to its complementary DNA or RNA than oligonucleotides or their analogs; resistance to enzymes like nucleases and proteases thereby showing long biological half-life; chemical stability over a wide pH range; no action as a primer; and an ability to act as gene promoters.

Improvements in genomic research have increased the rate of generation of information on the identity, structure and function of a number of human genes, thereby producing a diverse group of novel molecular targets for therapeutic and diagnostic applications. However, gene sequencing and characterization is still a slow and often arduous process, as evidenced by the fact that, to date, only about 3% of the entire human genome has been sequenced. The same advantageous binding and chemical stability properties that make PNAs useful as therapeutics and diagnostics also suggest such compounds will be useful in determining the sequence, structure and/or function of DNA and RNA.

In addition to completely characterizing a gene, unraveling the details of the interactions of the gene with its DNA binding proteins and determining the mechanisms whereby such proteins mediate gene expression, replication and transduction require a great deal of time and effort. Further, understanding the genetic malfunctions of dysfunctional genes that cause the many complex genetic disorders found in man still require extensive research. Thus, here too, PNAs can be useful.

While PNAs appear to be particularly well-suited for use as diagnostics, therapeutics and/or research tools, identification of appropriate PNAs for a specific purpose can be difficult, time consuming and expensive. For example, identifying which region of a gene should be targeted in order to provide a desired effect, such as blocking transcription thereof, or which region, if any, may be activated to promote transcription thereof, generally requires sequencing most, if not all, of the gene, then testing various PNA fragments complementary thereto.

Recently, combinatorial libraries of random-sequence oligonucleotides, polypeptides and/or synthetic oligomers have been employed to facilitate the isolation and identification of compounds capable of producing a desired biological effect or useful as diagnostics. Compounds so identified may mimic or block natural ligands, may interfere with the natural interactions of the target molecule or may simply be useful as tools for designing and developing other molecules with more desirable properties.

Combinatorial libraries useful in this general application may be formed by a variety of solution-phase or solid-phase methods in which mixtures of different subunits are added in a stepwise manner to growing oligomers, until a desired oligomer size is reached.

Alternatively, the library may be formed by solid-phase synthetic methods in which beads containing different sequence oligomers that form the library are alternately mixed and separated with one of a selected number of subunits being added to each group of separated beads at each step. An advantage of this method is that each bead contains only one oligomer species, allowing the beads themselves to be used for oligomer screening (Furka, *et al.*, Int. J. Pept. Protein Res. 37:487-493 (1991); Sebestyen, *et al.* Bioorg. Med. Chem. Letter 3:413-418 (1993).)

Still another approach that has been proposed involves the synthesis of a combinatorial library on spatially segregated arrays. (See, Fodor, *et al.*, Science, 251:767-773 (1991).) This approach has generally been limited in the number of different library sequences that can be generated.

Since the chance of finding useful ligands increases with the size of the combinatorial library, it is desirable to generate libraries composed of large numbers of different sequence oligomers. For example, in the case of oligonucleotides or oligonucleotide mimics, such as PNAs, a library having a 4-base variability and 8 oligomer residue positions (octamer) will contain 4^8 (65,536) different sequences to be a complete (universal) library. In the case of a 10 oligomer residue position (decamer) PNA or oligonucleotide universal library, 1,048,576 different sequences must be synthesized.

Because each different-sequence species in a large number library may be present in small amounts, one of the challenges in the combinatorial library screening procedure is isolating and determining the sequence(s) of species that have the desired binding or other selected properties. Thus, not only must the library be universal but the method(s) selected for screening that library

must be tailored to distinguish active from non-active species, considering the small amount of each species that is available.

Thus, what is needed is a system of improved techniques and tools for the rapid identification of agents, particularly PNAs, useful in characterizing genes, and useful as potential therapeutics and/or diagnostics. In particular, it would be desirable to have the advantages of specificity associated with a large species library, such as an octemeric or decameric library, without having to synthesize tens of thousands or millions of different-sequence species. It would be desirable to have rapid screening methods that may quickly identify "best candidates" from the library for further testing and/or development, and it would be desirable to have a means for determining optimal PNAs, both with respect to sequence base identity and length for use as therapeutics, diagnostics and/or research tools.

SUMMARY OF THE PRESENT INVENTION

The present invention addresses these and other needs by providing a system involving three key integrated components which permits the rapid identification and/or design of peptide nucleic acids (PNAs) capable of site-specific recognition of target nucleotide sequences and therefore useful as therapeutics, diagnostics and/or gene characterization tools. The first component of this system is a universal PNA library that may be easily and efficiently synthesized and that most preferably has the screening ability of a large library, such as an octameric library, yet does not require synthesis of a large number of individual species. The second component is a high throughput screening system, termed the Universal PNA

Identification (UPID™) System, that includes a number of assays designed to provide information on the binding activities of the different sequence PNAs to the target nucleotide sequence. The third component of the present invention is a software system especially designed to provide rapid analysis of the data collected from the (UPID™) System and to identify the sequence base identities and lengths of optimal PNA oligomers therefrom.

In one aspect the universal library incorporates universal nucleotide bases into each species in order to increase the screening capability of the library without the need for an unmanageable number of individual species. The universal library is subjected to the improved high throughput screening process to identify novel regulators acting by specific modulation of a selected gene, such as one implicated in human diseases. Optimization of these novel regulators is guided by the software system, thereby predicting the most appropriate therapeutic and/or diagnostic candidates, both in sequence length and sequence base identity. In addition, the present invention enables the structural/functional characterization of newly discovered genes and the identification of genomic mutations, such as single nucleotide polymorphisms, in either genomic DNA or PCR-amplified DNA for genetic diagnosis of disease states as well as the rapid screening of at-risk populations.

BRIEF DESCRIPTION OF DRAWINGS

The above and other aspects, features and advantages of the present invention will be more apparent from the following more particular description thereof, presented in conjunction with the following drawings wherein:

FIG. 1 is a schematic of a synthesis of an activated PNA monomer subunit according to the present invention;

FIG. 2 is a schematic of an alternative synthesis of an activated PNA monomer subunit according to the present invention;

FIG. 3 is a schematic of the synthesis of a resin-bound PNA trimer according to the present invention;

FIG. 4 is a schematic of an improved method of synthesis of a hexameric PNA oligomer according to the present invention;

FIG. 5 is a schematic of an alternative improved method of synthesis of a hexameric PNA oligomer; and

FIG. 6 is a schematic of the inverse synthesis of a PNA molecule according to the present invention.

DETAILED DESCRIPTION

While the universal PNA library contemplated herein may be constructed with any length of PNAs, for example, from tetramer to dodecamer, the longer the sequence of the PNAs the higher the specificity toward the complementary DNA or RNA gene fragments. However the synthesis of a library of PNA sequences of more than about 8 bases is not economically feasible, due to the large number of compounds that would need to be synthesized. For example, a dodecamer universal PNA library requires synthesis more than 16 million (4^{12}) discrete compounds. Thus, in one embodiment, an octameric library of PNAs is synthesized using standard solid-phase coupling and protection/deprotection reactions.

Even an octameric library, however, requires synthesis of numerous individual species, more than

65,000. Thus, in a preferred embodiment herein, one or more universal nucleotide bases are incorporated into the library, essentially as placeholders, thereby extending the library's usefulness without requiring synthesis of additional species. For example, in a most preferred embodiment, exemplified herein, a library having the sequencing capability of an octameric universal library, requires synthesis of only about 256 individual compounds. This is accomplished by incorporating, at every other nucleotide position in the octamer, a universal nucleotide base.

Current methodologies for the synthesis of peptide nucleic acids involve the stepwise addition of suitably protected PNA monomers via one of two standard synthesis protocols, depending on the protecting group strategy selected. This work has been described in detail in a number of recent papers, including: Dueholm, *et al.* J. Org. Chem., 59:5767 (1994); Thomson, *et al.*, Tetrahedron, 51:6179 (1995); Will, *et al.* Tetrahedron, 51:12069 (1995); Breipohl, *et al.*, Biorg. Med. Chem. Lett. 6:665 (1996); Koch, *et al.* J. Peptide Res., 49:80 (1997); Jordan, S., Bioorg. Med. Chem. Lett., 7:681 (1997); Breipohl, *et al.* Tetrahedron, 53:14671 (1997). In a typical procedure, the resin for solid phase synthesis is placed in the reaction vessel and treated with the activated acid of a first PNA monomer. The coupling reaction is performed, the reaction mixture is filtered or drained off; and the resin is washed and deprotected for the next synthesis step. In many cases a capping step is utilized to prevent the unreacted amino terminus from undergoing further chain extension. After washing the resin, the reaction sequence is repeated again carrying out this cycle until the desired full length oligomer mimic is obtained. Once the full length oligomers are synthesized, and if desired, they may be

further conjugated with detectable probes such as radioisotopes, fluorescence chromophores or stable metal isotopes bound to metallothionein, as is well known to those of skill in the art.

I. Improved Method of Synthesis of PNA Oligomers

Standard reactions may be employed in the synthesis of the universal PNA library contemplated herein, such as those disclosed in U.S. Patent Nos. 5,539,083 and 5,539,082, each of which is incorporated in its entirety by reference herein. However, an improved synthesis method, as described herein, is preferred. This improved method of synthesis provides enhanced recovery of unreacted starting material, and reduces the number of steps required to reach the desired sequence length. This method includes 1) use of inherently activated monomers subunits, thereby avoiding the need for activating agents in the coupling procedure, which activated monomers may be coupled to other monomers through a unique ring opening reaction; 2) a convergent synthesis involving coupling of PNA dimers or trimers to one another, thus, significantly reducing the number of steps required for sequence synthesis and improving purity and overall yields; and 3) an inverse synthesis protocol, enabling the synthesis of the PNA molecule from the N-terminus to the C-terminus by activation of the carboxylic acid on resin prior to the introduction of the next PNA monomer, thereby allowing for the recovery of unreacted PNA monomers in solution for recycling.

The synthesis of the activated monomer subunit is given in FIG. 1. Synthesis begins with the formation of the backbone subunit from ethylene diamine 10 and a bromoacetic acid or chloracetic acid equivalent 12 such as bromoacetic acid, bromomethyl acetate, chloroethyl acetate or bromoacetyl bromide. Slow addition of a

solution of the bromoacetic acid equivalent to a solution of ethylene diamine followed by heating gives the desired cyclic intermediate 14. Treatment of 14 with the activated nucleotide base ester 16 followed by protection of the resulting diamide gives the desired piperazinone 18. This cyclic intermediate 18 can be easily hydrolyzed to the N-Boc acid 20 to give a typical PNA monomer. Alternatively, the piperazinone intermediate 18 can be used as a pre-activated coupling reagent for addition of nucleophiles to yield a variety of products 22, 24, 26 and 28.

Turning to FIG. 2, alternatively and advantageously the preactivated, substituted PNA monomer subunits can be prepared by reductive amination of an amino aldehyde 40 or ketone 42 with the desired amino acid 44. Where R, R1 and R2 represent any aliphatic, aromatic, substituted aliphatic or substituted aromatic side chain. Subsequent deprotection and cyclization give the substituted piperazinone 46, which is converted to the pre-activated substituted PNA monomer 48 by addition of the activated nucleotide base ester 16. The cyclic intermediate 48 can be easily hydrolyzed to the N-Boc acid to give the free acid PNA monomer 50, or it may be utilized as a pre-activated coupling reagent for addition of nucleophiles to yield other products 52, 54, 56, and 58.

FIG. 3 provides a general scheme for the synthesis of a PNA trimer. Addition of the suitably protected preactivated PNA monomer 18 to a resin or other nucleophile followed by deprotection gives 70. Addition of a second PNA monomer 72 with subsequent deprotection gives 74. Repetition of this cycle continues until the desired chain length of PNA is reached, for example, a PNA trimer 76, as illustrated in FIG. 3. FIGS. 4 and 5 illustrate the convergent synthesis of a PNA molecule.

In this approach a library of suitably protected PNA dimers or trimers are used as the starting material for the synthesis. A PNA dimer 100 is coupled to a resin and deprotected to give 102. A second dimer is added by repeating the cycle to result in formation of the tetramer 104. The process is repeated again, and in only three coupling cycles, the hexamer 106 is prepared. This process significantly reduces the number of coupling steps and reduces the production of side products. Alternatively, and as shown in FIG. 5, suitably protected trimers can be used in this convergent synthesis approach thereby permitting formation of the PNA hexamer 106 in only two coupling steps. Similarly, both trimer and dimer libraries may be used in the convergent synthesis described herein to prepare PNAs of various lengths, for example, pentamer, heptamer, and undecamer, not amenable to preparation from either a dimer or trimer library alone.

The inverse synthesis of a PNA molecule is presented in FIG. 6. A suitably protected monomer 120 is coupled to a resin via an appropriate linker to give 122. Deprotection of the terminal acid, followed by activation and addition of a second monomer subunit gives 124. The coupling cycle can be repeated to give the trimer 126, and through continued cycling the desired PNA chain length is acquired. By activating the acid on the resin, this approach simplifies recovery of the PNA monomer thereby significantly reducing the cost of synthesis.

While the synthesis process just described has been detailed for a solid resin, those of skill in the art will appreciate that solution phase synthesis of PNAs may likewise be accomplished by simply replacing the resin with a suitable terminal functional group or side chain (Farese, A. *et al.* 37:1413, 1996).

II. Synthesis of Universal PNA Libraries

A. Use of Universal Nucleotide Bases

In a preferred embodiment contemplated herein, universal nucleotide bases, such as inosine, nitropyrole and/or 5-nitroindole, which are capable of binding to any of the four nucleotide bases, are incorporated into the universal library in specific, carefully chosen positions within each oligomer. Such universal-base-containing libraries require the synthesis of fewer individual PNA species, but have a comparable scope and function of their larger-sized counterparts. For example, in a most preferred embodiment, a universal library is prepared having at least four base positions and at least half the total number of base positions occupied by a conventional nucleotide base (A, C, T or G), with the remaining base positions being filled by a selected universal base.

Thus, for example, by substituting the conventional nucleotide bases located in just two positions in a nanomeric library - such as the third and seventh positions - with universal nucleotide bases, a universal library of only 16,384 (4^7) different PNA species need be synthesized. This library of "pseudo-heptamers" possesses comparable scope and function to the 262,144 species conventional nanomeric PNA library, yet is considerably more cost effective to prepare than its nanomeric counterpart.

In another embodiment, substitution of conventional nucleotide bases with universal bases may be expanded to involve every other nucleotide base position within the PNA. In this manner "mini-libraries" of particularly small size may be quickly, easily and inexpensively prepared, further enhancing the flexibility of universal libraries. For example, in a most preferred embodiment, an octameric library having only four conventional nucleotide bases and four universal

nucleotide bases is prepared. Such a universal library requires synthesis of only 256 different PNA sequences as opposed to the 65,536 PNA sequences required for a wholly conventional octameric PNA library. Again, the hybridization capacity of such a library is similar to that of the conventional counterpart. Further, the stability of the individual PNA species is greater in an octamer than in a tetramer. Thus, use of the universal octameric library containing universal nucleotide bases should provide significantly better results than use of a universal tetrameric library containing only conventional nucleotide bases.

While it is expected that increased substitution of conventional bases with universal nucleotide bases in a PNA library will reduce the usefulness of the library for certain purposes, such universal-base-containing libraries provide a fast, accurate, inexpensive tool for initial screenings of compounds. As will be discussed further below, the data and information collected from the screenings of such libraries may be used to determine the optimum sequence base identity of PNA oligomers for further screening and testing.

B. Apparatus for Library Synthesis - Hydrogel

Relatively large PNA libraries, according to the present invention, can be constructed using any one of a number of apparatus, including standard multiwell plates, microchips or spot membranes. For example, octameric PNA chips can be constructed by light-directed, spatially addressable chemical synthesis (Fodor, *et al.* Nature 364:555 (1993)). This approach makes use of a parallel process in which sites on a silica chip are illuminated through the apertures of a photolithographic mask, with resulting activation for chemical coupling with protected PNAs. Protection groups are removed from

terminal PNAs with illumination at each subsequent step, resulting in the activation of another set of sites for linkage with a new set of protected PNAs, *i.e.*, preferably the dimer or trimer selected for addition to the growing PNA. With this approach, PNAs are built up on the silica chip, with the location and composition of sites dependent on the pattern of illumination and the order of chemical coupling reagents used. The result is a high-density array in which the sequence of the PNA at each site is known. The binding of fluorescence-labeled DNA or RNA fragments to specific sites can be detected by fluorescence microscopy. This permits, for example, the identification of the DNA or RNA fragments based on the complementary PNA sequences to which they bind.

Although such a light-directed, spatially addressable chemical synthesis method may be employed, as well as various other apparatus and methods in the production of a PNA library according to the present invention, an improved method of library synthesis is presented herein. In this preferred embodiment, the PNAs are conjugated to a linker molecule having a free amine at the non-conjugated end. Using this amine, the PNAs are bound to a pre-hydrogel (such as Hypol, manufactured by Hampshire Chemical Corporation) having active isocyanate functionalities. In contrast to DNA-based oligonucleotides, PNAs are soluble in organic solvent and thus may be coupled to the pre-hydrogel using such organic solvents. The resulting PNA-hydrogel solution can be treated with a solution of water and water-miscible organic solvent, such as acetone. When the free isocynates react with water, the PNA-bound pre-hydrogel crosslinks and liberates CO_2. The addition of water-miscible organic solvent during the crosslinking process is used to control the amount and speed of CO_2

generation so that the resulting hydrogel remains transparent.

Most preferably, the universal PNA library according to the present invention is not randomly synthesized but is methodically synthesized. Methodic synthesis, as used herein, refers to synthesis of the library in such a manner that the location and sequence base identity of the individual species of the library are known, so that positive screening results do not require sequencing of the PNA oligomer. Thus, while the library synthesis may be and preferably is automatically performed, each species of the library is synthesized in a separate vessel, such as in separate microliter wells.

Once the PNA library, or the individual PNA oligomers to be tested, if less than the complete library, is isolated on the transparent hydrogel, screening the library with labeled DNA or RNA fragments is greatly facilitated. Thus, for example, the solvents are removed from the PNA-hydrogel membrane and the membrane is treated with the fluorescence-, radioisotope-, metal or otherwise-labeled DNA or RNA fragments in an aqueous medium. After washing the swollen hydrogel thoroughly to remove any unspecifically bound DNA or RNA fragments, the binding of labeled DNA or RNA fragments to specific sites can be monitored. To accelerate the washing procedure, electric fields can be applied with microelectrodes.

This process of creating a universal PNA library bound to a hydrogel membrane and then screening labeled compounds therewith, is simpler and less expensive than the microchip method and thus is preferable to that and other methods. Furthermore, the amount of PNA bound to the hydrogel can easily be controlled to provide a high density of PNAs; thus, a higher amount of labeled DNA or RNA fragments, can be

captured, thereby increasing the sensitivity of the assay. PNA-hydrogel formulations can be bound to microwell plates, or spotted onto membranes made of fabrics like nylon or similar polymers. Most preferably, which ever support is chosen, it will be transparent so that a multi-spot array assay, as described below, may easily be performed.

C. Immobilization of PNAs on Membrane Arrays

An alternative immobilization method, that is particularly useful for the rapid evaluation of DNA polymorphisms, is provided by screening PNA oligomers isolated on membrane arrays as described, for example, by Weiler J. *et al.*, Nuc Acid Res 25:2792, 1997. This method allows a large number of DNA samples to be quickly assayed. For example, arrays of up to 1,000 different PNA oligomers can be synthesized on a polymer membrane using a robotic device, such as the one developed by Frank, R. for parallel oligopeptide synthesis (Tetrahedron 48:9217, 1992). By using different enzymatically cleavable linkers to attach the PNAs to the membrane, individual PNA sequences can be selectively removed from the solid support for further analysis.

1. Attachment of Linkers to a Membrane Array

The attachment of linkers to an amino-functionalized membrane was accomplished by derivatizing the membrane with a peptide spacer, glutamic acid-(-tert-butylester)-(ϵ-aminohexanoic acid)-(ϵ-aminohexanoic acid) using standard Fmoc-chemistry. The respective amino acid derivative was activated by the addition of 1.2 equivalents diisopropylcarbodiimide (DIC) and 1 equivalent HOAt and used at a final concentration of 0.2 M in N-methyl-2-pyrrolidone (NMP). The membrane was submerged in this solution for 15 min., washed with dimethylformamide (DMF), and the Fmoc-groups were removed by a 5 min incubation in 20% piperidine in DMF. The

membrane was again washed in DMF, rinsed with ethanol and dried. Following addition of the spacer, the membrane was mounted in an Automated SPOT ROBOT (Frank, R. Tetrahedron 48:9217, 1992) and a grid of the desired format was spotted using, at each position, 0.3 μL of activated Fmoc-lysine-(ε-tert-butyloxycarbonyl). After a reaction time of 30 min, the membranes were treated with 5% acetic anhydride in dry NMP to cap all amino groups outside the spotted areas. The membrane was then washed and deprotected, and the spots were visualized with a solution of 0.01% bromophenol blue in DMF (Krchnak V. *et al.* Int. J. Pep. Protein Res., 32:415, 1988).

2. Conjugation of PNAs to the Membrane Array

PNA capture sequences were deposited to the individual spots on the membrane array using the ASP 222 Automated SPOT Robot. The complete synthesis cycle comprised: coupling (spotting of activated derivative followed by 20 min reaction time after placing material to the last spot); 5 min of acetylation in 5% acetic anhydride in DMF; five washes of 1 min each in 10 mL DMF; 5 min of deprotection with 20% piperidine in DMF; staining with 0.01% bromophenol blue in DMF; three rinses in 10 mL ethanol and subsequent drying. After completion of the synthesis, the PNA oligomers were deprotected by a 1 h incubation in a mixture of 90% trifluoroacetic acid, 5% water and 5% triethylsilane. Once immobilized on the array, the PNAs are ready for the high throughput screening process.

III. Improved High Throughout Screening Assays

Having synthesized the universal PNA library the next component to the present invention is screening the library for those PNA oligomers capable of binding to the target nucleotide molecule(s). Thus, in one aspect, the following assays are directed to the selection of

biologically active sequences as potential diagnostic and/or therapeutic candidates. Additionally, assays are provided that permit the structural and/or functional identification of newly discovered genes using PNA probes from the universal library. Finally, assays are provided that allow the identification of genomic mutations, such as single nucleotide polymorphisms, in either genomic DNA or PCR-amplified DNA for genetic diagnosis of disease states as well as the rapid screening of at-risk populations. Such screenings will be particularly useful for genetic counseling.

The assays described herein may be performed by hybridizing the PNAs to DNA fragments which are bound to solid supports or, alternatively, by immobilizing the PNAs and using them as DNA-capture agents to capture the test DNA sequence. Any of a number of methods of detecting PNA hybridization to the test DNA or RNA fragment.

A. DNA Hybridization with PNA Probes

1. Hybridization without Signal Amplification

One method contemplated herein for the screening of the PNA library, is hybridization of PNA probes to the DNA of interest without signal amplification. This method is particularly suited for the functional characterization of a newly discovered gene. The gene sequences of interest are immobilized on a solid support such as multi-well microplates or nylon membranes. Hybridization reactions are preferably performed in 10 mM sodium phosphate buffer, or equivalent, at pH 7.4 and at about 80°C, thereby taking advantage of the ability of PNAs to hybridize to their complementary DNA sequences under low salt and high temperature conditions. Non-specific binding sites are preferably blocked with a 1mg/mL solution of bovine serum

albumin (BSA), or a 0.1% SDS solution in 10 mM sodium phosphate buffer at pH 7.4 and washed thoroughly. The PNA probes are preferably labeled, such as with a radioactive isotope or biotin, prior to hybridization. Following hybridization of the labeled probe to the support-bound DNA, the label is activated. If the label is a radioisotope, of course, no activation is required. In the case of a biotin-labeled PNA probe, specific signal can be detected by reading biotin with avidin labeled with fluorescence emitters. Methods of fluorescence detection are well known to those of skill in the art. In general, the wavelength chosen to be measured by the fluorescence reader will be that corresponding to the particular type of fluorescence used to label the biotin, e.g., fluorescine, rhodamine, etc.

2. Hybridization with Signal Amplification

In some cases, for example, where the DNA sample of interest is only available in small quantities or where only a small portion of a lengthy strand of DNA is of interest, specific signal amplification may be required to enhance the signal-to-noise ratio and further increase the sensitivity of the assay. This is particularly relevant where PCR amplification of DNA is either not possible or not desirable. In a preferred method of signal amplification, the PNA probe is conjugated with a recognition tag which is then recognized by a second specific probe linked to biotin. This complex can then be detected with fluorescence-labeled avidin as previously described. The signal can be further amplified by adding multiple recognition steps similar to the multi-recognition steps currently used in multi-layer sandwich enzyme-linked immunoassays. As outlined above, hybridization most preferably takes place in 10 mM sodium phosphate buffer,

at pH 7.4 and at about 80°C. Following hybridization, the system is cooled to 37°C for signal amplification.

Examples of recognition tags useful for the signal amplification process as contemplated herein, include but are not limited to, bifunctional derivatives of macrocyclic tetrazacyclododecane chelates of the basic 13(ane)N4 formula, such as nitro-benzyl-DOTA (see, for example, Moi M.K. *et al.* J. Am. Chem. Soc. 110:6266, 1988; Renn O. and Meares C.F. Bioconj. Chem. 3:563, 1992; and U.S. Patent No. 4,678,667, 1987). Bifunctional DOTA derivatives contain four NH groups in a 12-atom ring configuration. The N-bound hydrogens are amenable to substitution with functional groups such as -CH2-COOH or -CH2-COOCH3; the nitrogens and the substituted carboxyl groups form coordination bonds with rare earth metal ions, including non-radioactive lanthanides like indium, gallium or lutetium, resulting in tight metal complexation.

A number of other bifunctional macrocyclic chelates may also be used herein, including PA-DOTA (alpha-[2-(4-aminophenyl)ethyl]-1,4,7,10-tetraazo-cyclodo decane-1,4,7,10-tetraacetic acid), and PA-DOTMA (alpha-[2-(4-amino-phenyl)ethyl]-1,4,7,10-tetraazo-cyclod odecane-1-acetic-4,7,10-tri(methylacetic) acid), as disclosed in US Patent No. 5,435,990 incorporated herein by reference, in its entirety. Still other bifunctional macrocyclic chelates of the basic 13(ane)N4 formula include TRITA, TETA and HETA (see, U.S. Patent No. 4,678,667, also incorporated herein in its entirety.) Additionally, linear chelates such as EDTA or DTPA may be used.

Another example of recognition tags that can be linked to PNA probes and used according to the present invention are DNA sequences capable of hybridizing

specifically with a second set of DNA or PNA probes which are themselves conjugated to the bifunctional macrocyclic chelates as described above. Such recognition tags must be carefully designed to prevent non-specific PNA-DNA hybridization with surface-bound DNA, which will mask the specific PNA-DNA hybridization and therefore interfere with the specificity of the assay. An example of a suitable probe as contemplated herein, is a bacteriophage lambda-specific sequence. DNA-bound PNAs containing such specific recognition tags can then be detected using the signal amplification systems described below.

The chelate recognition tags are preferably linked to each PNA or (DNA) probe via a space linker comprising, for example, an aliphatic chain containing 3 to 6 carbon atoms, to prevent the chelate from causing interference with the binding of the PNA probe to its complementary DNA sequence, such as by sterically hindering such binding. The DNA-bound PNA probe conjugated with the chelate recognition tag can be, and preferably is, detected using a specific anti-chelate monoclonal antibody (MoAb), such as a MoAb recognizing non-radioactive indium-labeled DOTA. The antibody is conjugated with biotin, and the DNA-bound PNA is then detected with fluorescence-labeled avidin.

Further signal amplification can be achieved by reacting the DOTA-conjugated PNA probe with a murine anti-DOTA MoAb. Such antibody can be probed with a second antibody, such as a goat anti-mouse antibody, which in turn can be probed with a third antibody, such as a sheep anti-goat antibody, which in turn can be probed with a fourth antibody, such as horse anti-sheep, and so forth. The last antibody probe, is biotin-conjugated prior to its addition to the reaction vessel and, following, its hybridization to the previous antibody, it is reacted with fluorescine-labeled avidin

and the resulting signal is detected as previously described.

In addition to using monoclonal antibodies to amplify the assay's signal, one of the antibodies, such as the horse-antisheep antibody outlined on the scheme above, can be probed with an antibody, such as a dog anti-horse, having a branched dextran of mw~40,000 conjugated thereto. This is then probed with an anti-dextran antibody which is conjugated to a branched, water-soluble polymer, such as polyvinyl alcohol or polyvinyl pyrrolidone, which polymer is further conjugated to biotin. The biotin is then reacted with fluorescence-labeled avidin as previously described and the significantly amplified signal is detected as above.

Advantageously, the method of screening DNA fragments with PNA probes conjugated to a recognition tag can be modified to permit the consecutive detection of multiple DNA sequences in the same well. In this modified method different PNA probes are labeled with different chelates which can then be recognized by their corresponding MoAbs. Each antibody is then probed with avidin conjugated to a different fluorescence-emitters. In this way each probe may be separately detected by the characteristic wavelength of light it emits.

B. Alternative Detection Methods Using PNA Probes

One alternative detection system contemplated herein, is the use of luminescence rather than fluorescence. According to this method, the antibody used in the final recognition step is conjugated with an enzyme such as alkaline phosphatase instead of biotin. The signal is then developed by addition of a luminescent phosphatase substrate and detected using a standard luminometer.

Another alternative detection system contemplated herein is based upon detection of the

kinetics of association (ka) or dissociation (kd) of the DNA-PNA hybrid using, for example, a BIAcore (Biomolecular Interaction Analysis) surface-plasmon resonance detection apparatus (Pharmacia, Uppsala). According to this method, a gold-covered sensor chip within the detection apparatus is treated with a layer of dextran linked to avidin. A PNA probe, conjugated via a linker to biotin, is immobilized on the avidin-dextran coated surface and captures complementary DNA sequences. The signal measured by the apparatus is proportional to the change of the refractive index at the surface of the chip and is generally assumed to be proportional to the mass of the substance bound to the chip (Jonsson 1991; Karlsson 1993). By measuring the amount of bound substance as a function of time, when a solution containing the complementary strand passes over the chip, the kinetics of association of the DNA to the immobilized PNA probe can be determined. This system functions only in a single read-out mode, and thus, may not be suitable for analysis of large samples. However, it may be suitable when detailed kinetic determinations of specific PNA-DNA interactions are desired or required.

C. Polymorphism Detection Using PNA Probes

In a modified detection assay, contemplated herein, PNAs having sequences complementary to conserved sequences adjacent to areas of DNA containing single nucleotide polymorphisms or other areas of interest are immobilized on a solids support. The immobilized PNAs are then exposed to the DNA fragments of interest and specific PNA-DNA hybridization is allowed to occur. After removing any non-hybridized DNA, the PNA-bound DNA is detected using one of the above-described methods. By selectively capturing only DNA fragments containing the mutation of interest, this modified assay permits

concentration of the DNA sample, further increasing the signal to noise ratio.

D. Biological Activity Assay

The biological activity of selected PNAs may initially be assessed in cell free systems by various methods. For example, direct inhibition of protein synthesis may be tested using cellular extracts capable of carrying out protein synthesis upon addition of m-RNA. In such an assay, PNAs are added to the cellular extracts in addition to the m-RNAs to be translated. By comparing the translation products of the extracts containing PNAs to control assays without PNAs, inhibitory PNA sequences can be identified. Similarly, inhibition of m-RNA transcription can also be assessed using extracts that will synthesize specific m-RNA in response to added DNA from the gene of interest. The PNAs are added to the extracts and the m-RNA product therefrom is compared to the RNA product from extracts having only the DNA and no PNA added thereto. Alternatively, activation of gene transcription, rather than inhibition, can be assessed using the same system described above but measuring increased levels of m-RNA, instead of decreased levels, in the cell-free transcription system. Alternatively, activation of gene transcription can be assessed with transcription-active extracts. These extracts include genes coding for appropriate transcription repressors or attenuators. The DNA fragment of interest is added to the reaction. If addition of a particular PNA blocks inhibition of transcription, by for example, binding to the DNA encoding the transcription inhibitor, an <u>increase</u> in m-RNA production will result. Active sequences, identified by any of these methods, can then be analyzed, by the improved software system described below, to determine the optimum PNA sequence base identity and

sequence length for affecting the target nucleotide sequence (generally, a DNA or RNA sequence).

IV. Method for Optimizing Base Sequence and Sequence Length

According to the present invention a method for designing the most appropriate PNA oligomer(s) for affecting the target nucleotide sequence by optimizing the sequence base identity and sequence length of the PNA(s) is provided. The method uses an algorithm to manipulate data concerning the relationships between binding constants, structures and/or sequences of specific PNAs to predict optimal base sequence identities and sequence lengths of PNAs for use as potential therapeutic and/or diagnostic candidates. In addition, the software system contemplated herein can perform the rapid sequencing of newly discovered or poorly characterized genes. Additionally, where the sequence of a target gene is known, the algorithms contemplated herein can easily and quickly identify all possible complementary PNAs from the universal library, that should be tested for biological activity. This of course, greatly simplifies the screening process by dramatically reducing the number of individual PNA sequences that must be tested.

Preferably, when a double strand DNA with a known sequence is targeted, the sequences of both parallel and antiparallel strands are screened and tested, thereby doubling the effort in screening, while resulting in two sets of data for further analysis by the new algorithms provided herein. Ideally, of course, PNA candidates derived from the parallel strand should complement those derived from the antiparallel strand. However, to the degree such does not occur, for example, due to the inherent limitations of the testing methods,

the software system contemplated herein will appropriately adapt its analysis to include relevant variable data while excluding spurious results. In operation, a set of data from the high throughput screening of the individual PNA oligomers is evaluated to determine an ideal sequence base identity and sequence length for binding to the target molecule and affecting that target's function as desired. First, an active hit is identified (for example, B4 - B11 of FIG. 7, wherein "B#" indicates the sequential location of the particular base). Next, additional PNA sequences from the library which overlap the sequence of the initial active hit, are identified and are ordered in single base shifts to the left or to the right of the initial hit. Data obtained from testing of the sequences identify the quality and quantity of activity for each sequence. As shown in the example in FIG. 7, the first overlapping sequence to the left of the B4-B11 sequence is B3-B10 and is active whereas the second overlapping sequence, B2-B9, is inactive, suggesting the left boundary for the optimal active PNA sequence is B3. Similarly, moving to the right, the sequence B5-B12 is active whereas the sequence B6-B13 is inactive, suggesting the right boundary for the optimal PNA sequence is B12. In this example then, the optimum PNA sequence is the decamer, B3-B12.

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	Target
			B4	B5	B6	B7	B8	B9	B10	B11				Initial Hit
		B3	B4	B5	B6	B7	B8	B9	B10					Active
	B2	B3	B4	B5	B6	B7	B8	B9						Inactive
				B5	B6	B7	B8	B9	B10	B11	B12			Active
					B6	B7	B8	B9	B10	B11	B12	B13		Inactive

Further contemplated herein is use of the software system to trace a pattern of hybridization of labeled gene fragments to a PNA universal library in order to determine the sequence of the gene of interest. For example, a DNA fragment having an unknown sequence is

labeled with a detectable probe and then hybridized with individual PNAs which are positioned on a support in a specific rule-based and traceable manner. Upon hybridization with a first PNA oligomer, software searches for related PNA sequences to determine to what degree the gene fragment binds thereto. This process continues until a full-length sequence for the fragment is determined. Performing this assay, simultaneously for numerous fragments from a single gene permits the rapid sequencing of those fragments and, therefore, of the gene.

While the invention herein disclosed has been described by means of specific embodiments and applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the scope of the invention as set forth in the claims.

CLAIMS

What is claimed is:

1. An improved peptide nucleic acid universal library comprising at least one universal nucleotide base located internally in each oligomer species within the library.

2. The peptide nucleic acid universal library of Claim 1, comprising at least four conventional nucleotide bases and at least four universal nucleotide bases in each oligomer species within the library.

3. The peptide nucleic acid universal library of Claim 2, wherein the conventional bases and universal bases of each oligomer species are located in alternate positions with respect to one another.

4. The peptide nucleic acid universal library of Claim 3, wherein the universal bases are selected from the group consisting of inosine, nitropyrole and 5-nitroindole.

5. The peptide nucleic acid universal library of either of Claim 2 or Claim 3, wherein each oligomer species within the library comprises at least twelve nucleotide bases.

6. A method of designing a peptide nucleic acid (PNA) capable of site-specific recognition of a target nucleotide sequence comprising the steps of:

a) providing a universal peptide nucleic acid library wherein at least four base positions in each PNA oligomer are occupied by conventional nucleotide bases;

b) exposing the target nucleotide sequence, under binding conditions, to a set of PNA oligomers from the universal library;

c) identifying the sequence base identities of the PNA oligomers from the set that bind to the target nucleotide sequence;

d) analyzing the activities of the PNA oligomers that bind to the target nucleotide sequence;

e) comparing the activity data from step (d) and the sequence base identities from step (c); and

f) determining an optimal sequence length and sequence base identity of at least one PNA oligomer capable of site-specific recognition of the target nucleotide sequence based upon the comparisons made in step (e).

7. The method of Claim 6 wherein the step of providing a universal peptide nucleic acid library further comprises the steps of:

a) preparing a universal dimer library of peptide nucleic acid oligomers;

b) preparing a universal trimer library of peptide nucleic acid oligomers; and

c) coupling the peptide nucleic acid oligomers of the dimer and trimer libraries to generate a larger universal peptide nucleic acid library having at least four nucleotide base positions in each oligomer.

8. The method of Claim 6, wherein the step of providing a universal peptide nucleic acid library further comprises providing a universal peptide nucleic acid library wherein at least one base position in each PNA oligomer is occupied by a universal nucleotide base.

9. The method of either Claim 6 or Claim 8, wherein the steps of exposing the target nucleotide sequence to a set of PNA oligomers from the universal library and identifying the sequence base identities of the PNA oligomers from the set that bind to the target nucleotide sequence further comprise the steps of:

a) covalently binding either (i) the PNA oligomers of the universal library or (ii) the target nucleotide sequence to a polymeric resin;

b) labeling the other of (i) and (ii) to constitute a detectable sequence;

c) exposing the resin-bound sequence, under binding conditions, to the detectable sequence;

d) washing the resin-bound sequence to remove unbound detectable sequence; and

e) detecting the label of the detectable sequence to determine whether and to what extent detectable sequence is bound to the resin-bound sequence.

10. The method of Claim 9, wherein the covalent binding step comprises binding the PNA oligomers of the universal library to a hydrogel.

11. The method of Claim 6, wherein the step of determining an optimal sequence length and sequence base identity of at least one PNA oligomer capable of site-specific recognition of the target nucleotide sequence is preformed in software.

12. An improved method for the preparation of a universal library of peptide nucleic acids of at least four nucleobases in length, the method comprising the steps of:

a) providing a universal library of resin-bound peptide nucleic acid dimers;

b) providing a universal library of peptide nucleic acid dimers that are not resin bound, wherein one base in each species of the non-resin-bound library is activated for coupling to another base; and

c) coupling the activated peptide nucleic acid dimers to the resin-bound dimers to form a universal library of at least four nucleobases in length.

13. The method of Claim 12, wherein the step of providing a universal library of resin-bound peptide nucleic acid dimers comprises the steps of:

a) providing activated, protected peptide nucleic acid monomer subunits;

b) coupling a first activated, protected peptide nucleic acid monomer subunit to the resin;

c) deprotecting the coupled first peptide nucleic acid monomer;

d) coupling a second activated, protected peptide nucleic acid monomer to the coupled first peptide nucleic acid monomer;

e) deprotecting the second peptide nucleic acid monomer, thereby resulting in a resin-bound peptide nucleic acid dimer; and

f) repeating steps (b)-(e) to form a universal library of resin-bound peptide nucleic acid dimers.

14. The method of Claim 13, wherein the step of providing activated, protected peptide nucleic acid monomer subunits comprises the steps of:

a) providing a solution of ethylene diamine;

b) slow addition of a solution of bromoacetic acid equivalent to the solution of ethylene diamine;

c) heating of the mixed solution to form a cyclic intermediate;

d) addition of an activated nucleotide base ester resulting in a diamide; and

e) protection of the diamide, resulting in a piperazinone that may be readily hydrolyzed to a peptide nucleic acid monomer.

15. A method of optimizing the sequence base identity and sequence length of peptide nucleic acid oligomers to improve the site-specific recognition capabilities thereof, the method comprising the steps of:

a) providing a peptide nucleic acid universal library, wherein each oligomer species within the library comprises at least eight peptide nucleic acid bases and wherein at least two of each peptide nucleic acid bases is a universal nucleotide base;

b) selecting a number of the oligomer species from the peptide nucleic acid universal library for testing;

c) testing the binding characteristics of the selected number of the oligomer species with the target nucleotide sequence; and

d) analyzing the binding characteristics of the selected number of oligomer species to determine optimum sequence base identity and sequence length.

16. The method of Claim 15 wherein the step of analyzing the binding characteristics of the selected number of oligomer species to determine optimum sequence base identity and sequence length is performed in software.

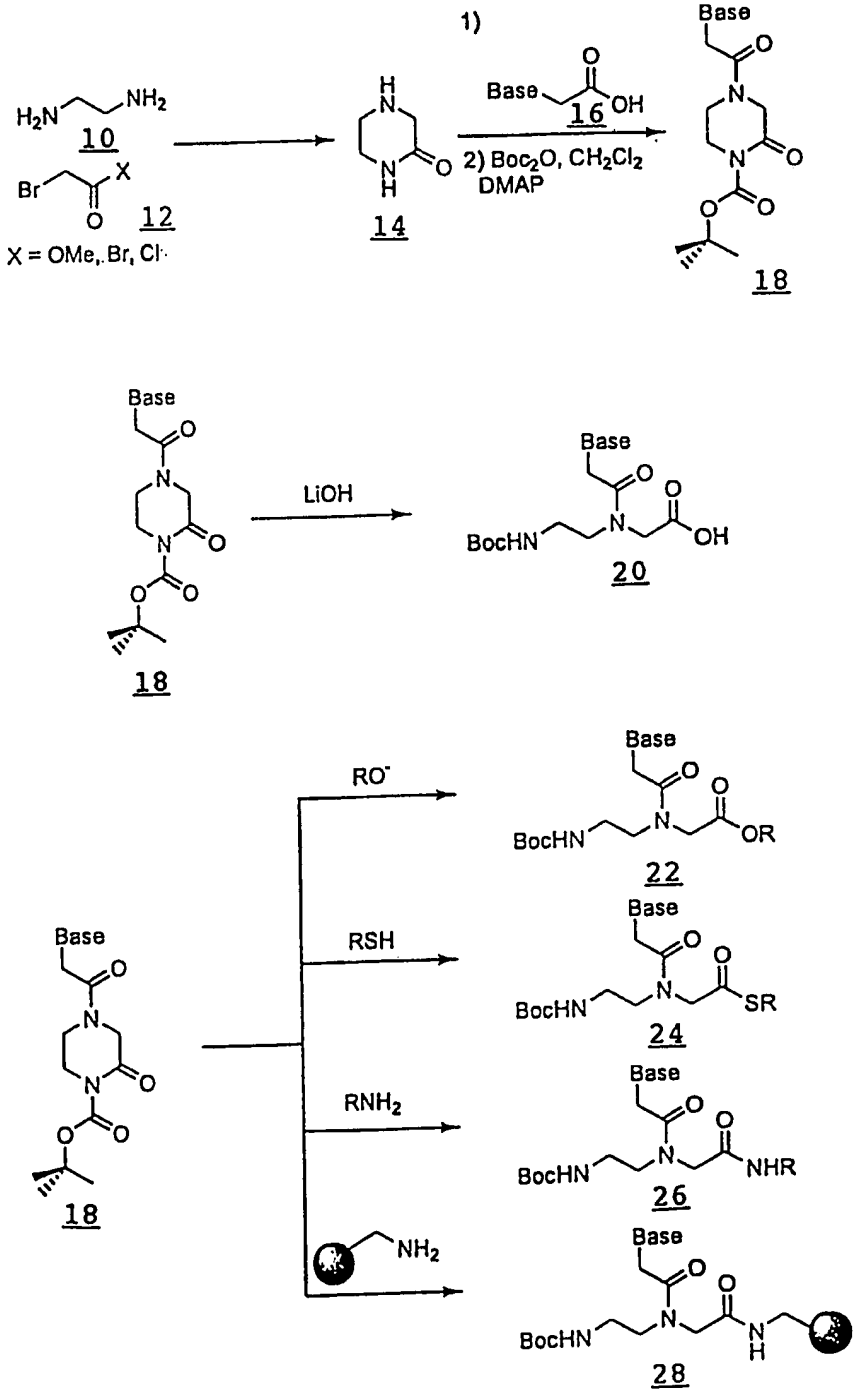

1)
Base
16
OH
2) Boc2O, CH2Cl2
DMAP
H2N
NH2
10
Br
X
O
12
X = OMe, Br, Cl
14
18
LiOH
BocHN
20
RO-
RSH
RNH2
NH2
22
24
26
28
OR
SR
NHR
NH

FIGURE 1

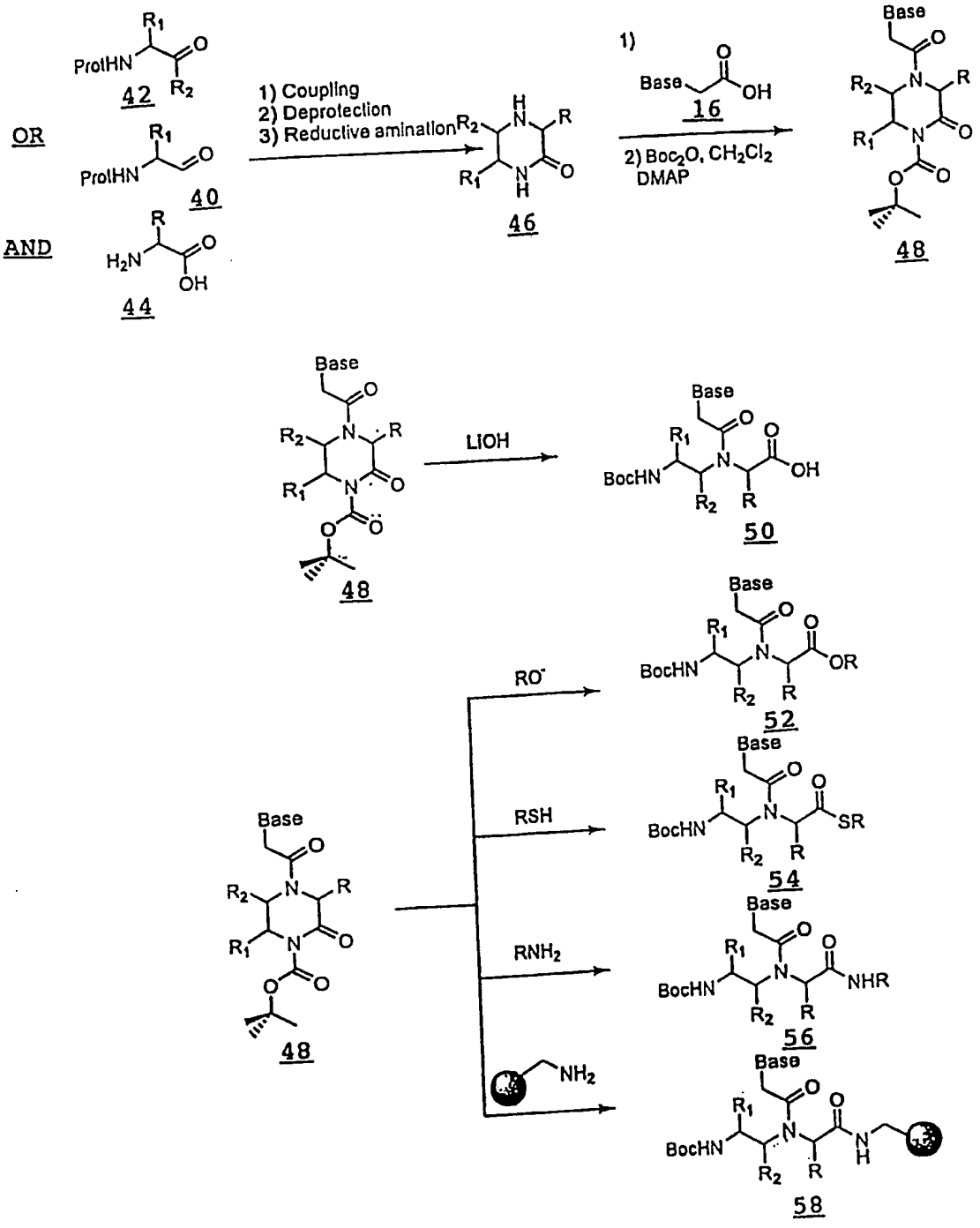

ProtHN
42
OR
ProtHN
40
AND
H_2N
OH
44
1) Coupling
2) Deprotection
3) Reductive amination
46
1)
Base
OH
16
2) Boc_2O, CH_2Cl_2
DMAP
48
LIOH
BocHN
50
RO⁻
52
RSH
SR
54
RNH_2
NHR
56
NH_2
58

FIGURE 2

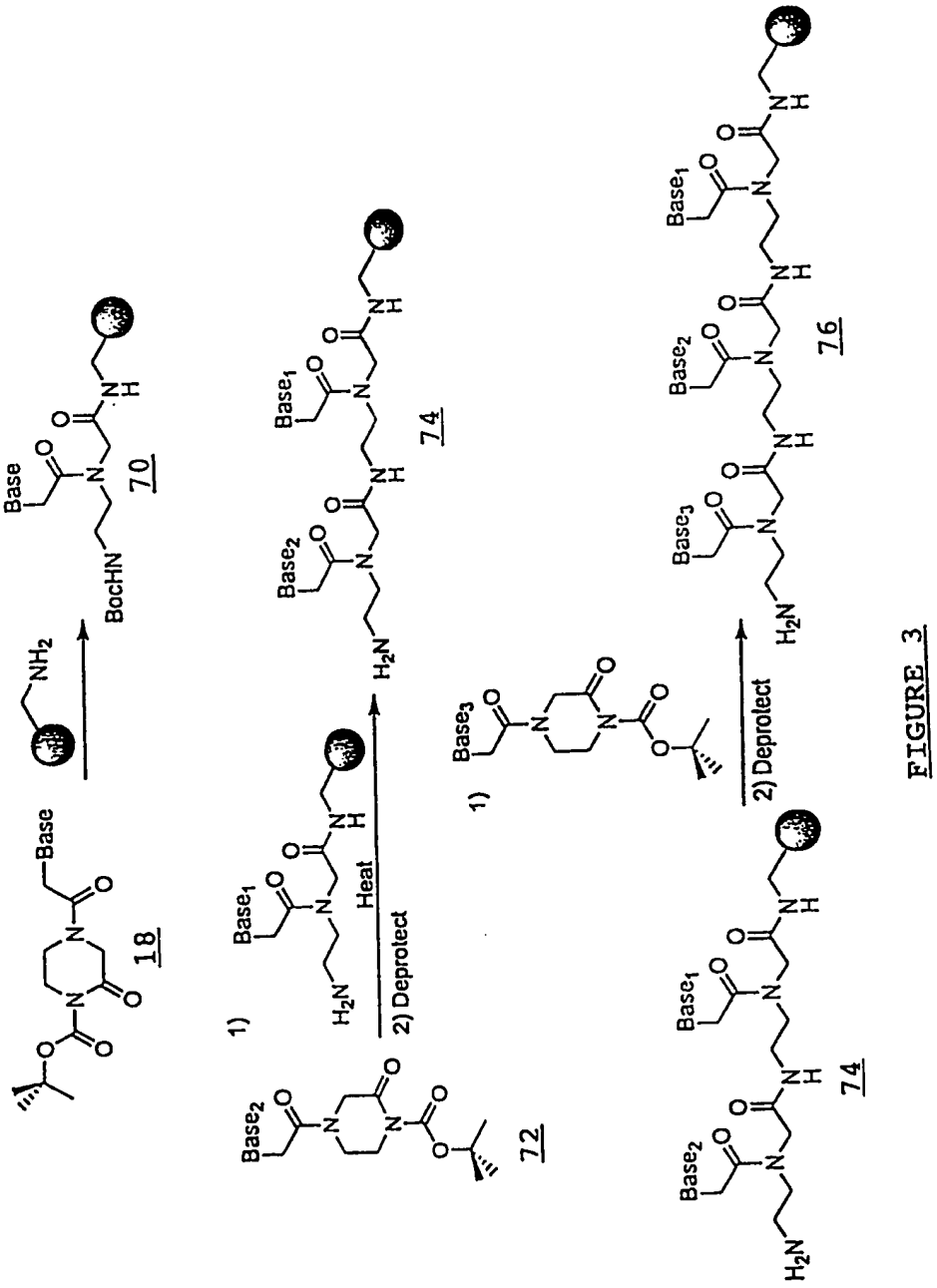

Base
NH2
BocHN
70
18
1)
Heat
2) Deprotect
Base1
Base2
H2N
74
72
Base3
76

FIGURE 3

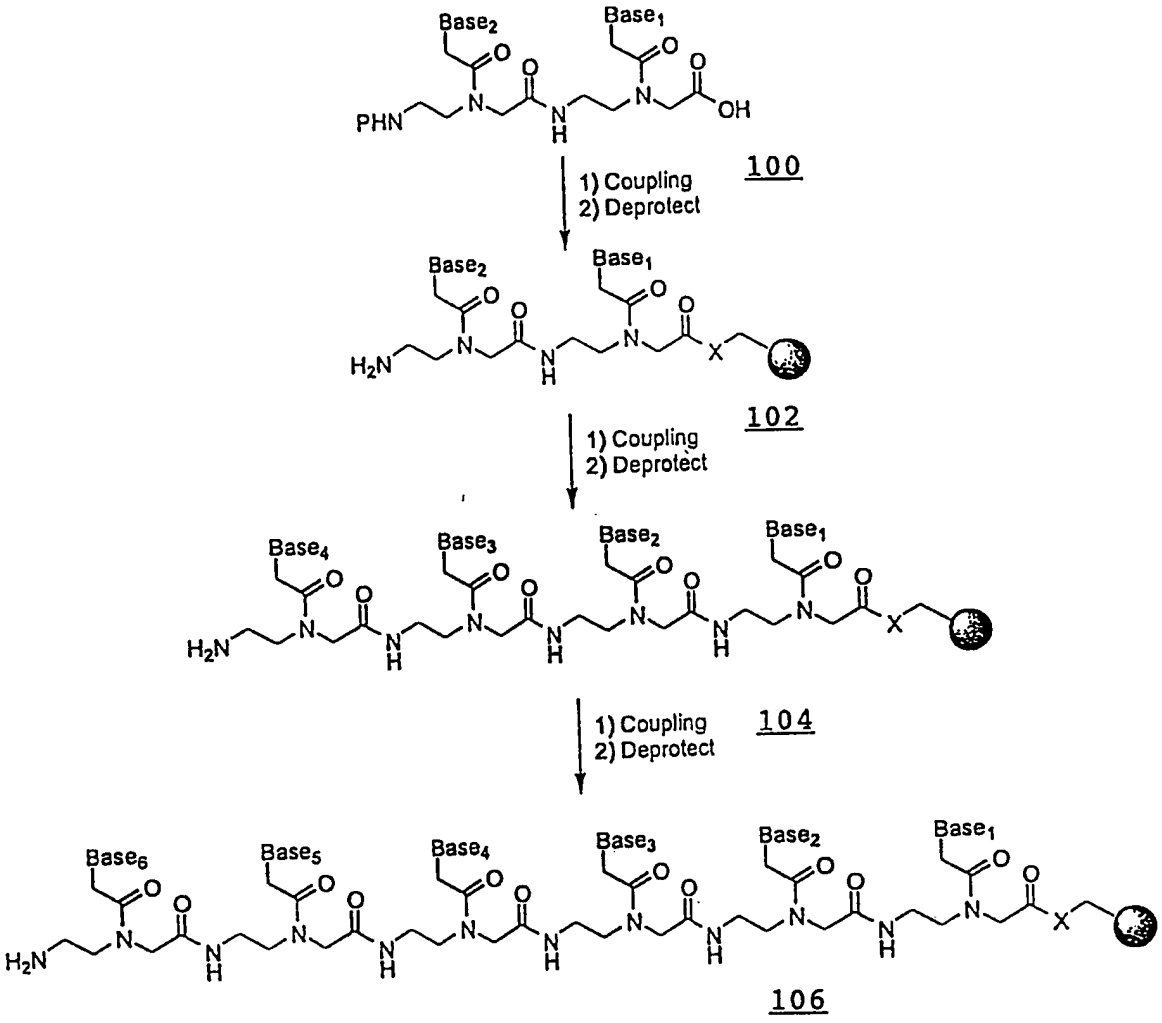
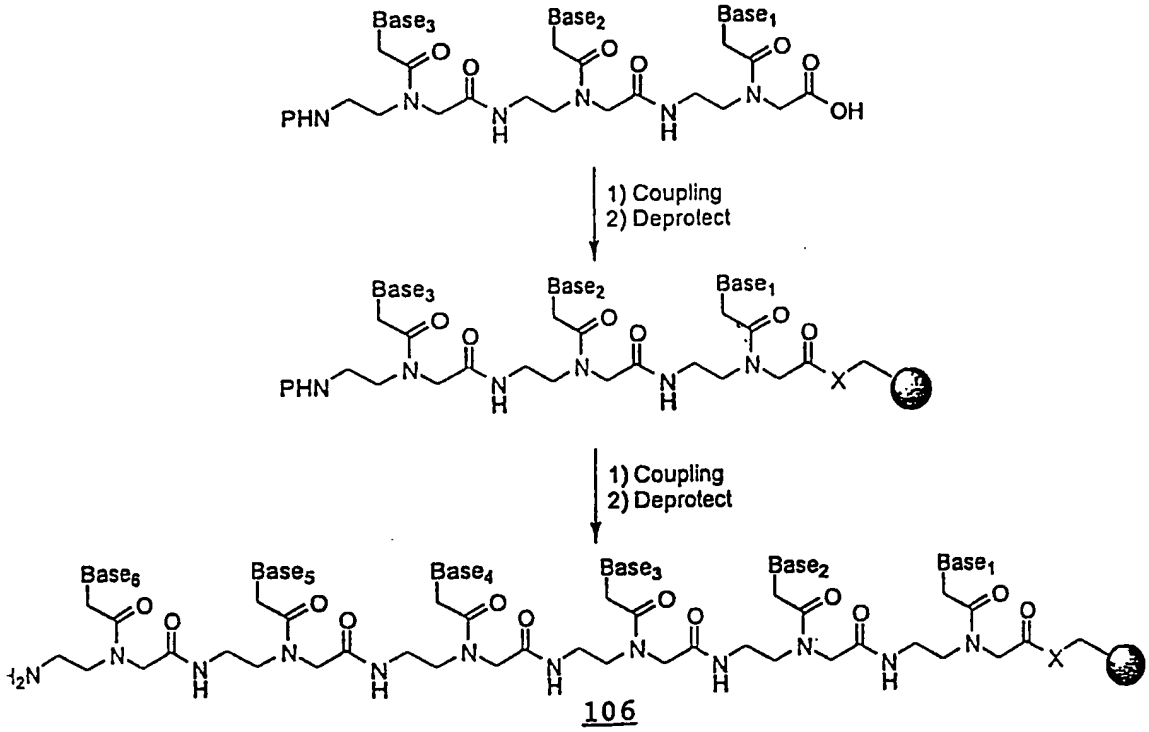

Base2
Base1
PHN
OH
100
1) Coupling
2) Deprotect
H2N
X
102
Base4
Base3
104
Base6
Base5

FIGURE 4


Base3
Base2
Base1
PHN
OH
1) Coupling
2) Deprotect
X
Base6
Base5
Base4
H2N

106

FIGURE 5

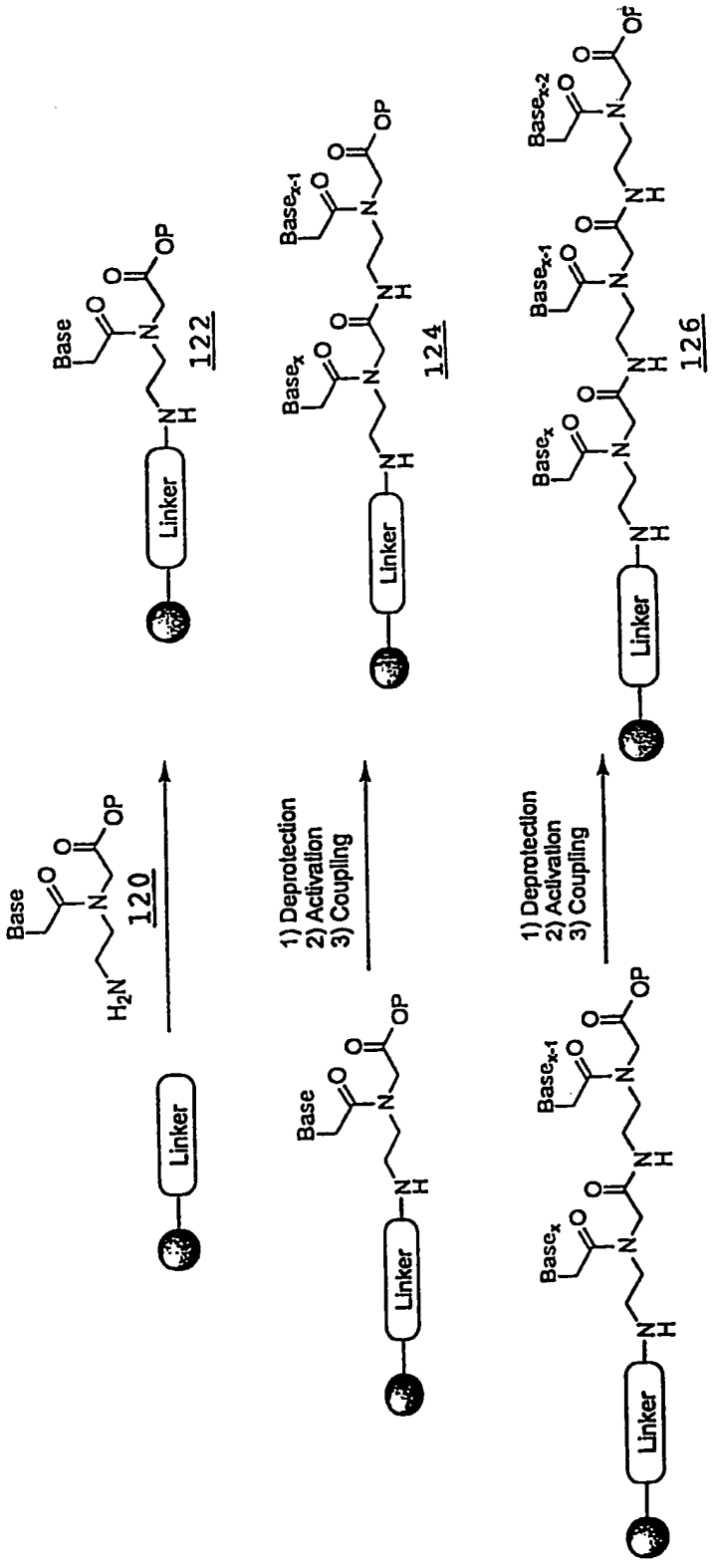

Base
120
H_2N
Linker
1) Deprotection
2) Activation
3) Coupling
122
124
126
OP
$Base_x$
$Base_{x-1}$
$Base_{x-2}$

FIGURE 6

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 99/14969

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K1/04 C07K14/00 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95 23163 A (ISIS) 31 August 1995 (1995-08-31) the whole document	1-16
Y	WO 95 14706 A (ISIS) 1 June 1995 (1995-06-01) the whole document	1-16
Y	DE 195 32 553 A (HOECHST AG) 6 March 1997 (1997-03-06) the whole document	1-16
Y	WO 96 20212 A (D BUCHARDT ET AL.) 4 July 1996 (1996-07-04) the whole document	1-16

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 October 1999	04/11/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Masturzo, P

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

...formation on patent family members

Inter ...nal Application No
PCT/US 99/14969

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9523163 A	31-08-1995	US 5539083 A	23-07-1996
		AT 185572 T	15-10-1999
		AU 684152 B	04-12-1997
		AU 1926195 A	11-09-1995
		CA 2183371 A	31-08-1995
		EP 0777678 A	11-06-1997
		JP 11209393 A	03-08-1999
		JP 9503523 T	08-04-1997
		US 5864010 A	26-01-1999
		US 5831014 A	03-11-1998
WO 9514706 86 A		NONE	
DE 19532553 A	06-03-1997	AU 708034 B	29-07-1999
		AU 6440896 A	06-03-1997
		CA 2184681 A	05-03-1997
		EP 0761681 A	12-03-1997
		JP 9124572 A	13-05-1997
		NO 963677 A	05-03-1997
		US 5817811 A	06-10-1998
WO 9620212 A	04-07-1996	AU 4729796 A	19-07-1996